Eckley, IA., & Nason, G. (2018). A test for the absence of aliasing or local white noise in locally stationary wavelet time series. *Biometrika*, *105*(4), 833-848. https://doi.org/10.1093/biomet/asy040, https://doi.org/10.1093/biomet/asy040

Biometrika (2018), **00**, 0, *pp.* 1–16
Printed in Great Britain
doi: 10.1093/biomet/asy040

A test for the absence of aliasing or local white noise in locally stationary wavelet time series

BY I. A. ECKLEY

Department of Mathematics and Statistics, Lancaster University, Bailrigg, Lancaster LA1 4YF, U.K.

i.eckley@lancaster.ac.uk

AND G. P. NASON

School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, U.K.

g.p.nason@bristol.ac.uk

SUMMARY

Aliasing is often overlooked in time series analysis but can seriously distort the spectrum, the autocovariance and their estimates. We show that dyadic subsampling of a locally stationary wavelet process, which can cause aliasing, results in a process that is the sum of asymptotic white noise and another locally stationary wavelet process with a modified spectrum. We develop a test for the absence of aliasing in a locally stationary wavelet series at a fixed location, and illustrate its application on simulated data and a wind energy time series. A useful by-product is a new test for local white noise. The tests are robust with respect to model misspecification in that the analysis and synthesis wavelets do not need to be identical. Hence, in principle, the tests work irrespective of which wavelet is used to analyse the time series, although in practice there is a trade-off between increasing statistical power and time localization of the test.

Some key words: Aliasing; Local spectrum; Sample rate; Subsampling; White noise; Wind energy.

1. INTRODUCTION

Typically a data analyst is presented with a time series sampled at a fixed rate. However, the series might have been sampled at a higher rate, or the analyst could request future samples at a higher rate. In practice, it is important to question whether time series are sampled often enough to successfully capture their second-order structure. Improper sampling can lead to aliasing, which this article proposes to detect and locate via a new test.

For a time series sampled at intervals of length Δ, the range of angular frequencies in the spectrum that can be observed undistorted is $[0, \pi/\Delta)$, where π/Δ is the Nyquist frequency (Chatfield, 2003, p. 109). If the highest frequencies in a series exceed the Nyquist frequency, then aliasing occurs and distorts the spectrum estimable by any method. As the spectrum and autocovariance are Fourier duals, distortion of the spectrum implies distortion of the autocovariance, and techniques that rely on either can be affected. Consequently, aliasing can have a critical impact.

Without knowledge of the data-generation process, the analyst will, a priori, be unaware of whether aliasing has occurred, and could misguidedly analyse the series assuming it is free from distortion. For example, if $\{X_t\}$ is a real-valued stationary process with spectrum $f_X(\omega)$ for

$\omega \in [0, \pi)$, and $Y_t = X_{2t}$ is observed, then the spectrum of Y_t is $f_Y(\omega) = f_X(\omega) + f_X(\pi - \omega)$ for $\omega \in [0, \pi/2)$. Generally, f_X cannot be identified from f_Y, so estimation of f_X is impossible from an estimate of f_Y. Thus, there is a need for principled approaches to testing for the presence of aliasing.

Hannan (1960), Priestley (1983), Hamilton (1994), Bloomfield (2000), Brillinger (2001) and Chatfield (2003) all describe aliasing, but provide no advice on how to detect or locate it. Instead, they suggest steps that can be taken to guard against it. One possibility is to apply a low-pass anti-aliasing filter prior to sampling, to ensure that the highest frequency in the filtered series is below the Nyquist rate. Although anti-alias filtering can be useful for analogue or very high sample-rate signal acquisitions, it is not useful when data are acquired at slow rates, such as in economics. Moreover, even if anti-alias filtering is used, valuable high-frequency information could be lost. Another possibility is that one might know, a priori, the highest frequency contained within a time series, and can choose the sample rate high enough to prevent aliasing.

Hinich & Wolinsky (1988) and Hinich & Messer (1995) introduced a hypothesis test for aliasing in stationary time series based on the bispectrum, a third-order quantity. Our wavelet test is based on simpler second-order quantities, and is designed for nonstationary time series. Computing the bispectrum typically takes $O(T^2)$ operations. Our test is faster, requiring only $O(T \log T)$ operations, which can be important for long time series.

Previous work has considered alias detection at a fixed location for a nonstationary time series, where the true signal or spectrum is known beforehand. Wunsch & Gunn (2003) used ice core time series and induced aliasing to demonstrate how it can lead to misleading scientific conclusions. Our test does not require knowledge of the underlying true spectrum or the higher-rate time series.

In a single realization, a nonstationary series can sometimes be aliased and sometimes not, depending on its spectral content relative to the Nyquist frequency at a given point. Hence, with a nonstationary series one may ask not only whether the series is aliased, but also where. Our test is designed to help answer both questions for locally stationary wavelet time series. That wavelets bring something genuinely new to the aliasing problem can be seen by replicating our method using Fourier-based quantities. Unlike the wavelet equations, which have a solution as given below, the equivalent Fourier-based equations are underdetermined, which is precisely the usual aliasing problem.

If one has prior knowledge that the time series has been properly sampled and there is no possibility of aliasing, then our test becomes a test for local white noise. Although global white noise tests are popular and useful, at the time of writing we are unaware of any local method tailored to locally stationary series. Local tests can be used for similar purposes to global ones, such as assisting with model selection, understanding time-varying forecasting performance or, as suggested by a referee, detecting measurement error in some circumstances.

2. REVIEW OF LOCALLY STATIONARY WAVELET PROCESSES

Locally stationary wavelet processes are time series models, constructed from wavelets, that change their statistical properties slowly over time. They are particularly useful for their ability to model time series operating at dominant scales in areas such as finance (Fryzlewicz, 2005), economics (Winkelmann, 2016), ocean engineering (Killick et al., 2013), structural engineering (Spanos & Kougioumtzoglou, 2012), energy (Nowotarski et al., 2013; de Menezes et al., 2016) and business (Michis, 2009). Dahlhaus (2012) provides a comprehensive review of locally stationary time series. We now briefly review essential definitions from Nason et al. (2000).

DEFINITION 1 (Discrete wavelets). *Let $\{h_k\}$ and $\{g_k\}$ be the low- and high-pass quadrature mirror filters underlying the Daubechies (1992) compactly supported orthogonal continuous-time wavelets. The discrete wavelets $\psi_j = (\psi_{j,0}, \psi_{j,1}, \ldots, \psi_{j,N_j-1})$ are vectors of length N_j for scales $j \in \mathbb{N}$ obtained using the formulae*

$$\psi_{1,n} = \sum_k g_{n-2k}\delta_{0,k} = g_n \quad (n = 0, \ldots, N_1 - 1), \tag{1}$$

$$\psi_{j+1,n} = \sum_k h_{n-2k}\psi_{j,k} \quad (n = 0, \ldots, N_{j+1} - 1), \tag{2}$$

where $N_j = (2^j - 1)(N_h - 1) + 1$, with N_h being the number of non-zero elements of $\{h_k\}$, and $\delta_{0,k}$ is the Kronecker delta. The number of vanishing moments of the associated continuous-time Daubechies compactly supported wavelet is $N = N_h/2$, where $\int x^m \psi(x)\,\mathrm{d}x = 0$ for $m \in \mathbb{N}$ such that $0 \leqslant m < N$. Such wavelets are commonly referred to as Daubechies DN *wavelets.*

DEFINITION 2. *A locally stationary wavelet process is a sequence of doubly indexed stochastic processes $\{X_{t,T}\}_{t=0,\ldots,T-1}$ ($T = 2^J, J \in \mathbb{N}$) having the following representation in the mean-square sense:*

$$X_{t,T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k;T}\,\psi_{j,k-t}\,\xi_{j,k}, \tag{3}$$

where $\{\xi_{j,k}\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ is a collection of uncorrelated random variables with zero mean and unit variance, $\{\psi_{j,k}\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ is a set of discrete wavelets, and $\{w_{j,k;T}\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ is a set of amplitudes satisfying the following conditions. For each $j \in \mathbb{N}$, there exists a Lipschitz-continuous function $W_j : (0,1) \to \mathbb{R}$ such that: (i) $\sum_{j=1}^{\infty} |W_j(z)|^2 < \infty$ *uniformly in $z \in (0,1)$;* (ii) *the Lipschitz constants, L_j, are uniformly bounded in j and $\sum_{j=1}^{\infty} 2^j L_j < \infty$;* (iii) *there exists $\{C_j\}_{j\in\mathbb{N}}$ such that for each T, $\sup_k |w_{j,k;T} - W_j(k/T)| \leqslant C_j/T$, where for each j the supremum is over $k = 0, \ldots, T-1$ and $\{C_j\}$ is such that $\sum_{j=1}^{\infty} C_j < \infty$.*

Spectral power for a locally stationary wavelet time series is quantified by the evolutionary wavelet spectrum, the time-scale analogue of the usual stationary spectrum, $f(\omega)$.

DEFINITION 3. *The locally stationary wavelet process $\{X_{t,T}\}_{t=0,\ldots,T-1}$, for $T \geqslant 1$, has evolutionary wavelet spectrum defined by $S_j(z) = |W_j(z)|^2$ for $j \in \mathbb{N}$ and $z \in (0,1)$ with respect to $\{\psi_{j,k}\}$.*

Hence, the $\{w_{j,k}\}$ are a collection of amplitudes such that $w_{j,k}^2 \approx S_j(k/T)$. Evolution of the second-order properties of $X_{t,T}$ is controlled by smoothness constraints on $S_j(z)$ as a function of z via those imposed on $W_j(z)$ in Definition 2(i)–(iii). For brevity, we henceforth drop the second subscript, T, in $X_{t,T}$.

The spectrum, $S_j(z)$, governs the contribution to variance in X_t at different scales at time z. Informally, $S_j(z)$ corresponds to the process variance integrated over the approximate frequency band $[2^{-j}\pi, 2^{1-j}\pi]$. For example, the approximate band for $S_1(z)$ is $[\pi/2, \pi]$, that for $S_2(z)$ is $[\pi/4, \pi/2]$, and so on. The precise frequency bands and their degree of overlap depend on the particular wavelet $\psi_{j,k}(t)$ used in (3). We next recall several key quantities associated with locally stationary wavelet theory.

DEFINITION 4. *The raw wavelet periodogram of Y_t is defined to be $I_{\ell,m} = d^2_{\ell,m}$ for $\ell \in \mathbb{N}$ and $m \in \mathbb{Z}$, where $\{d_{\ell,m}\}$ are the nondecimated wavelet coefficients of Y_t given by $d_{\ell,m} = \sum_t Y_t \psi_{\ell,m}(t)$ for $\ell \in \mathbb{N}$ and $m \in \mathbb{Z}$. The autocorrelation wavelet is given by $\Psi_j(\tau) = \sum_k \psi_{j,k}\psi_{j,k-\tau}$ for $j \in \mathbb{N}$ and $\tau \in \mathbb{Z}$, and the inner product operator of the autocorrelation wavelets is $A_{j,\ell} = \langle \Psi_j, \Psi_\ell \rangle = \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau)$ for $j, \ell \in \mathbb{N}$.*

Further information on locally stationary wavelet processes can be found in Fryzlewicz & Nason (2006), Van Bellegem & von Sachs (2008) and Nason (2008, Chapter 5).

3. LOCALLY STATIONARY WAVELET MODELS UNDER DYADIC SAMPLING

3·1. *Dyadic subsampling of locally stationary wavelet processes*

In this article, aliasing is induced by starting with a locally stationary wavelet process, X_t, and then forming $Y_t = X_{2^r t}$ ($t \in \mathbb{Z}; r = 1, \ldots, J-1$) by dyadic subsampling of X_t. Since Y_t is sampled at a slower rate than X_t, high-frequency power in X_t could reappear as aliased power in Y_t. Our first result shows that Y_t can be represented by the sum of a locally stationary wavelet process, constructed using the same wavelets as the original, and another process, which is asymptotically white noise.

THEOREM 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a locally stationary wavelet process with evolutionary wavelet spectrum $\{S_j(z)\}_{j=1}^\infty$. If $Y_t = X_{2t}$, then $\{Y_t\}_{t\in\mathbb{Z}}$ admits the decomposition $Y_t = L_t + F_t$, where L_t is a locally stationary wavelet process with the same underlying wavelets as X_t and having raw wavelet periodogram expectation given by the right-hand side of (4), and F_t is a process with mean zero and autocovariance $\operatorname{cov}(F_t, F_{t+\tau}) = S_1(2t/T)\delta_{0,\tau} + O(T^{-1})$. Hence, F_t is asymptotically white noise with variance $S_1(z)$. Further, if $S_1(z)$ is constant for $z \in (0,1)$, then F_t is stationary white noise with variance S_1.*

The proof is given in the Appendix. Theorem 1 can be extended to repeated dyadic sampling as follows.

COROLLARY 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be as in Theorem 1, and let $Y_t = X_{2^r t}$ ($r \in \mathbb{N}$). Then, asymptotically, $\{Y_t\}$ admits the decomposition $Y_t = L_t + F_t$, where L_t is a locally stationary wavelet process with the same underlying wavelets as X_t and having raw wavelet periodogram expectation as in the right-hand side of (4), and F_t is a process with mean zero and autocovariance $\operatorname{cov}(F_t, F_{t+\tau}) = \delta_{0,\tau} \sum_{j=1}^r S_j(2^r t/T) + O(T^{-1})$. Further, if $S_1(z), \ldots, S_r(z)$ are all constant functions of $z \in (0,1)$, then F_t is stationary white noise with variance $\sum_{j=1}^r S_j$.*

Nason et al. (2000) developed an estimator for the wavelet spectrum by exploiting the raw wavelet periodogram $d^2_{\ell,m}$ of X_t using the result $E(d^2_{\ell,m}) = \sum_j A_{\ell,j} S_j(m/T) + O(T^{-1})$. The next result explains what happens to $E(d^2_{\ell,m})$ after dyadic subsampling.

THEOREM 2. *Suppose that $\{X_t\}_{t\in\mathbb{Z}}$ is a locally stationary wavelet process with evolutionary wavelet spectrum given by $\{S_j(z)\}_{j=1}^\infty$ with Daubechies compactly supported wavelets. The expectation of the raw wavelet periodogram $d^2_{\ell,m}$ of $Y_t = X_{2^r t}$ is*

$$E(d^2_{\ell,m}) = \sum_{j=1}^{r} S_j(2^r m/T) + \sum_{j=r+1}^{\infty} A_{j-r,\ell} S_j(2^r m/T) + O(T^{-1}) \quad (r = 1, \ldots, 2^{J-1}), \qquad (4)$$

where $\ell \in \mathbb{N}$, $m \in \mathbb{Z}$ and A is the inner-product operator from Definition 4. The result also holds for Shannon wavelets if the evolutionary wavelet spectrum, $S_j(z)$, has continuous first derivative for each $j > 0$.

Theorem 2 generalizes Proposition 4 of Nason et al. (2000), in which $r = 0$, and additionally establishes the result for Shannon wavelets. The Shannon wavelet can be thought of as the limiting case of Daubechies wavelets with an infinite number of vanishing moments. A consequence is that $A_{j,\ell} = 2^j \delta_{j,\ell}$ for Shannon wavelets; see Nason et al. (2000) for further details.

Aliasing results in power redistribution across scales. For example, with $r = 1$ and subsampling dyadically once, formula (4) for Y_t becomes

$$E(d^2_{\ell,m}) = S_1(2m/T) + \sum_{j=2}^{\infty} A_{j-1,\ell} S_j(2m/T) + O(T^{-1}). \tag{5}$$

Compare (5) with the usual formula for the asymptotic expectation of the raw wavelet periodogram of X_t without subsampling, due to Nason et al. (2000):

$$E(d^2_{\ell,m}) = \sum_{j=1}^{\infty} A_{j,\ell} S_j(m/T) + O(T^{-1}). \tag{6}$$

There are three key differences between (5) and (6). First, $E(d^2_{\ell,m})$ in (5) is contaminated by $S_1(2m/T)$ at every analysis scale, i.e., for all $\ell \geqslant 1$. This contamination is the manifestation of aliasing in the locally stationary wavelet domain and could be used to detect aliasing, as described in §3·2. The second difference is that the mixing matrix on the right-hand side of (5) is $A_{j-1,\ell}$, not $A_{j,\ell}$. The third difference is that the dyadically subsampled periodogram exists on the grid $2m/T$ rather than on m/T.

3·2. *White noise confounding and hypothesis specification*

If a locally stationary wavelet process X_t is white noise with variance σ^2, then the unsampled process is such that $E(d^2_{\ell,m}) = \sigma^2$ for $\ell \in \mathbb{N}$ and $m \in \mathbb{Z}$; see the proof of Lemma B.3 in Fryzlewicz et al. (2003). The same quantity appears at every scale as in the first term of (4) or (5) when subsampling. Hence, the effects of white noise and aliasing are confounded in our set-up.

Such confounding is not a surprise, as white noise can be seen as the ultimate aliased signal. For example, suppose that X_t is a stationary process with variance $\sigma^2 < \infty$ and autocovariance $\gamma_X(\tau) \to 0$ as $\tau \to \infty$. If $Y_t = X_{2^r t}$, then $\gamma_Y(\tau) \to \sigma^2 \delta_{\tau,0}$ as $r \to \infty$. In other words, repeated subsampling of X_t leads to white noise. Taking a broader view, in a practical situation, what appears to be white noise may be the result of repeatedly subsampling a time series. Similar confounding can be found in the test for aliasing for stationary series in Hinich & Messer (1995); rejection of the null hypothesis there can mean that the series is not random, not stationary, aliased, not mixing, or any combination of these.

However, even with confounding, we can still test the null hypothesis H_0 of there being no white noise component and no aliased component at $z_0 \in (0, 1)$ against the hypothesis H_A that a white noise or aliased component exists at z_0. So although we cannot separate the two components, our hypothesis test can test locally whether they are both absent.

4. ALIASING/WHITE NOISE TEST

4·1. *The test procedure*

Let $t_0 \in \{1, \ldots, T\}$ and let $z_0 = t_0/T$ be the rescaled-time equivalent of t_0. If we could observe the evolutionary wavelet spectrum $\{S_j^{(Y)}(z_0)\}_{j\in\mathbb{N}}$ directly, it would be possible to test the ideal null hypothesis $H_0^{(\mathrm{I})}$ that there exists $j^* \in \mathbb{N}$ such that $S_{j^*}(z_0) = 0$, versus the alternative $H_{\mathrm{A}}^{(\mathrm{I})}: S_j(z_0) > 0$ for all $j \in \mathbb{N}$. In view of §3·2, the null hypothesis means that $\{Y_t\}$ cannot have any additive white noise component; and, in view of Theorem 2, it would also mean that no aliasing has occurred from any underlying subsampled $\{X_t\}$ process.

However, the spectrum $\{S_j^{(Y)}(z)\}_{j\in\mathbb{N}}$ cannot be observed directly but can only be estimated from a realization $\{Y_t\}_{t=1}^T$ on a finite set of scales $j \leqslant J^\dagger < J$, where $J^\dagger$ is chosen to avoid boundary effects. Hence, we can only gain information on $S_j(z_0)$ for $j \in \{1, \ldots, J^\dagger\}$, and therefore only test the null hypothesis H_0 that there exists $j^* \in \{1, \ldots, J^\dagger\}$ such that $S_j^*(z_0) = 0$, versus the alternative $H_{\mathrm{A}}: S_j(z_0) > 0$ for $j \in \{1 \ldots, J^\dagger\}$. Clearly, if H_0 is true, then $H_0^{(\mathrm{I})}$ is true. However, if H_{A} is true, then either $H_0^{(\mathrm{I})}$ or $H_{\mathrm{A}}^{(\mathrm{I})}$ could be true, as $S_j(z_0)$ might be zero for scales $j > J^\dagger$ that cannot be discerned from a finite-length series.

We proceed by testing whether $S_j(z_0) = 0$ for each scale $j \in \{1, \ldots, J^\dagger\}$ separately. Let $b \in \mathbb{N}, b > 0$ be a window width, and define $R_b = \{-b, \ldots, b\}$. Define the sample $\{\hat{S}_{j,\,t_0+r}\}_{j=1,\ldots,J^\dagger,\,r\in R_b}$, where $\hat{S}_{j,\ell}$ is the estimator given by $\sum_{i=1}^{J^\dagger} A_{j,i}^{-1} I_{i,\ell}$ with $I_{i,\ell}$ and A as in Definition 4. Proposition 4 of Nason et al. (2000) and the Lipschitz continuity of $S_j(z)$ imply that $E(\hat{S}_{j,\,t_0+r}) = S_j(z_0) + O(r/T)$. Hence, under H_0, there exists $j^* \in \{1, \ldots, J^\dagger\}$ such that $E(\hat{S}_{j^*,\,t_0+r}) = O(r/T)$. For small r and large T we have $E(\hat{S}_{j^*,\,t_0+r}) \approx 0$, with the $\hat{S}_{j^*,\,t_0+r}$ having the same distribution on $r \in R_b$ asymptotically. This approximation improves with increasing T and for smoother S around z_0. Hence, we employ standard tests of zero location on the set $\{\hat{S}_{j,\,t_0+r}\}_{r\in R_b}$, one for each scale $j = 1, \ldots, J^\dagger$. For example, we could use Student's t-test on the sample $\{\hat{S}_{j,\,t_0+r}\}_{r\in R_b}$. Let $\bar{S}_{j,\,t_0}$ and $\hat{\sigma}_{j,\,t_0}^2$ be the sample mean and variance of $\{\hat{S}_{j,\,t_0+r}\}_{j=1,\ldots,J^\dagger,\,r\in R_b}$. The next result establishes the asymptotic distribution of the usual t-statistic operating on the $\{\hat{S}_{j,\,t_0+r}\}_{j=1,\ldots,J^\dagger,\,r\in R_b}$ sample.

THEOREM 3. *Let $j \in \{1, \ldots, J^\dagger\}$ be fixed, and let $b \in \mathbb{N}, b > 0$. Suppose that $\{Y_t\}$ is a stationary Shannon wavelet process with innovations that satisfy $E(|\xi_{j,k}|^6) < \infty$. Let $t = (2b)^{1/2}\bar{S}_{j,\,t_0}\hat{\sigma}_{j,\,t_0}^{-1}$ be Student's t-statistic defined on the sample $\{\hat{S}_{j,\,t_0+r}\}_{j=1,\ldots,J^\dagger,\,r\in R_b}$. Then, as $b \to \infty$, t converges in distribution to a normal variable with mean $2^j S_{j,\,t_0}$ and variance* 1.

We use Student's t-test or, in the case of significant violations of normality, a nonparametric test such as the Wilcoxon signed rank or the signmedian test. Student's t-test does not work well in the presence of autocorrelation, which is known to exist in the $\hat{S}_{j,\,t_0+r}$ sequence as a function of r; see Jones (1975), for example. To improve the performance of the t-test, we employ the equivalent sample size method from Zwiers & von Storch (1995), which uses estimated autocorrelation to alter the effective sample size. The t-test is robust against violations of the normality assumption; see the Supplementary Material.

We use Holm's method (Holm, 1979) to control the overall size of our test over the multiple $J^\dagger$ hypotheses. This is a reasonable choice, as no assumption is made about correlations between the $\hat{S}_j(z_0)$ at different scales, but as a result our test is somewhat conservative.

4·2. *Practical considerations*

Our test is robust with respect to mismatch of the synthesis wavelet in (3) and the analysis wavelet given in Definition 4, because a white noise component and aliasing caused by subsampling both cause a constant to be added to each level of the wavelet periodogram, regardless of the type of wavelet used to synthesize the process. Our test uses a wavelet method to detect whether a constant has been added to each level, and it works irrespective of the analysis wavelet.

All Daubechies compactly supported wavelet spectral estimates suffer from spectral leakage, where power from one scale can leak to adjacent scales. Such leakage makes it harder to determine whether power in a given scale is zero, and this influences the statistical size and power of our test. In practice, there is a trade-off between mitigating spectral leakage by using longer, smoother wavelets (Nason et al., 2000, § 4.1) and achieving good time localization by using shorter, rougher wavelets such as Haar wavelets. Therefore, we recommend using Daubechies D5, D6 or D7 mid-range wavelets, which achieve a good compromise.

Similarly, the size and power of our test will depend on the window width, b, which should be large enough to enable detection of modest departures from H_0, but small enough to ensure that the estimates $\hat{S}_{j,\,t_0+r}$ are representative of the behaviour of the spectrum at z_0. An appropriate upper bound for b could be obtained by computing an estimator of the spectrum (Nason et al., 2000; Fryzlewicz & Nason, 2006) and then choosing b such that the window of spectral estimates on R_b is approximately constant for each scale. For fixed T our test operates over a small interval $(z_0 - b/T,\, z_0 + b/T)$, although, conceptually, as T increases this interval will shrink to z_0.

Torrence & Compo (1998) introduced the wavelet cone of influence as the region of the wavelet spectrum degraded by edge effects. The extent of the cone depends on the length T of the time series and the length N_h of the wavelet filter used to compute the wavelet spectrum from Definition 1. The coarsest non-cone scale is the largest $J_{\mathrm{NC}} \in \mathbb{N}$ such that $J_{\mathrm{NC}} \leqslant \log_2\{(T/2 + N_h - 2)/(N_h - 1)\}$. Our test uses scales $1, \ldots, J^\dagger$ and, to avoid edge effects, we choose $J^\dagger \leqslant J_{\mathrm{NC}}$. For a functioning practical test, with reasonable power, we recommend $T \geqslant 512$, and for the range of sample sizes considered we choose $J^\dagger = 4$. This provides us with enough scales to furnish a nontrivial test, but not too many scales, which would reduce power and expose the test to large autocorrelations at the coarser scales, as mentioned earlier. For practical use, we recommend that $J^\dagger$ grow linearly with $\log T$.

We already employ multiple hypothesis testing methods to control the overall size of our test by combining the results over $J^\dagger$ scales using Holm's method. Similar methods could be used if we repeated our test at multiple time locations, not just at a single time-point z_0.

4·3. *Post hoc investigation if the test is rejected*

If the null hypothesis is rejected, then evidence exists that aliasing or a white noise component is present, and we describe three post hoc strategies to distinguish them.

One strategy might be to perform stationarity tests on a range of times around z_0 and, if stationary, use the Hinich & Wolinsky (1988) or Hinich & Messer (1995) aliasing tests. Another strategy could be to reanalyse the series at a faster rate, if possible. If significant power exists at finer scales, then aliasing probably has occurred, and the new rate should be considered in future.

A third approach, illustrated in §5·2, investigates the local spectral properties of the series in clear patches, where there is no aliasing or white noise, to discover whether power moves from low to higher frequencies just before an aliasing/white noise event or vice versa. This exploratory approach has the advantage of not requiring faster-sampled data required by the second approach or the stationarity tests of the first approach, which could have poor power for moderate sample sizes.

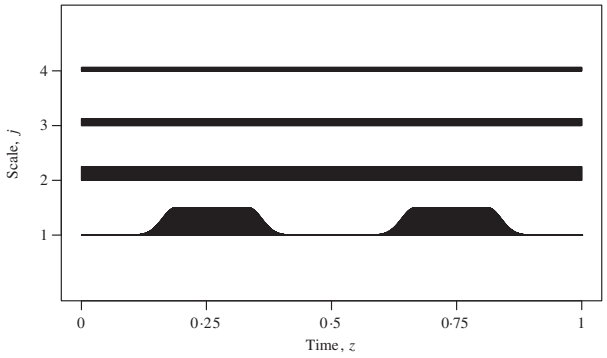

Fig. 1. The test-bed spectrum $\{S_j^{\mathrm{TB1}}(z)\}_{j=1}^{J^\dagger=4}$.

Table 1. *Empirical rejection rate* (%)*, per scale and overall, at* $z_0 = 1/2$ *with no subsampling*

b	Size $j=1$	Power $j=2$	Power $j=3$	Power $j=4$	Size Overall
16	3·2	42·1	28·0	30·9	0·6
32	2·5	76·2	36·4	39·4	1·1
64	1·8	99·2	69·7	65·6	1·2
128	2·2	100·0	98·0	90·0	2·1

For the specific case of Shannon locally stationary wavelet processes, Eckley & Nason (2014) showed that aliased power can be detected and the aliasing effects removed.

5. SIMULATIONS AND A WIND ENERGY EXAMPLE

5·1. *Simulation study*

Our simulation study uses the test-bed evolutionary wavelet spectrum $\{S_j^{\mathrm{TB1}}(z)\}$ shown in Fig. 1 to produce realizations without subsampling, $r = 0$. The test-bed spectrum satisfies the null hypothesis for $J^\dagger = 4$ at $z_0 = 1/2$, and the alternative at $z_0 = 1/4$.

We drew 1000 realizations of length $T = 4096$ from the locally stationary process model of (3) using the test-bed spectrum and Daubechies D5 wavelets with normally distributed innovations. Tables 1–4 show the empirical rejection rates; in each table, the rightmost column gives the rejection rate obtained by combining the per-scale p-values via Holm's method, whereas the rates for individual scales $j = 1, \ldots, 4$ are reported in the preceding columns. Using the method described in §4·1, we test $H_0 : S_j(z_0) = 0$ versus $H_{\mathrm{A}} : S_j(z_0) > 0$, for each scale $j = 1, \ldots, J^\dagger$ separately, at a nominal size of 5%, using the equivalent sample size version of Student's t-test.

Table 1 illustrates our test working as a local white noise test, with no subsampling. Under the null hypothesis, at $z_0 = 1/2$, the test shows increasing empirical power at scales $j = 2, 3, 4$ as the window width b increases, and has empirical size ranging from 1·8% to 3·2% for scale $j = 1$. These empirical sizes are somewhat conservative, and we attribute this to negative correlations present in the $\{\hat{S}_{j,\, t_0+r}\}_{r \in R_b}$ sample. The equivalent sample size method mentioned in §4·1 attempts to improve the size calibration, but the outcome is not perfect.

The overall empirical size of our test is 1–2%, lower than the nominal rate. Under the alternative hypothesis, at $z_0 = 1/4$, Table 2 shows the increasing empirical power for all scales.

Table 2. *Empirical power* (%) *of test, per scale and overall, at* $z_0 = 1/4$ *with no subsampling*

b	$j = 1$	$j = 2$	$j = 3$	$j = 4$	Overall
16	76·7	28·6	25·9	31·4	3·2
32	98·4	55·2	40·4	43·6	11·7
64	100·0	94·6	77·6	73·0	54·6
80	100·0	98·8	86·6	83·0	70·6
96	100·0	99·4	94·1	89·1	83·3
128	100·0	99·8	98·4	94·0	92·3

Table 3. *Empirical rejection rate* (%)*, per scale and overall, at* $z_0 = 1/2$ *with dyadic subsampling* ($r = 1$)

		Power	Size	Power	Size
b	$j = 1$	$j = 2$	$j = 3$	$j = 4$	Overall
16	–	87·7	9·7	63·5	2·8
32	–	95·5	5·8	88·6	1·3
64	–	100·0	5·9	99·8	2·2
128	–	100·0	2·4	100·0	0·9

Table 4. *Empirical rejection rate* (%)*, per scale and overall, at* $z_0 = 1/4$ *with subsampling* ($r = 1$)

b	$j = 1$	$j = 2$	$j = 3$	$j = 4$	Overall
16	–	78·6	16·6	38·7	2·1
32	–	98·7	29·9	60·7	6·5
64	–	100·0	62·4	90·6	29·0
128	–	100·0	94·3	100·0	65·8

The next simulation set illustrates application of our absence-of-aliasing test using a second test-bed spectrum, $S_j^{\mathrm{TB2}}(z)$, which is identical to $S_j^{\mathrm{TB1}}(z)$ except that the scale $j = 3$ has zero power. This set subsamples using $r = 1$, which reduces the realization length from 4096 to 2048 according to §3·1, and induces aliasing where power existed at scale $j = 1$. The subsampling causes scale $j = 1$ to disappear, and so we test the individual scale hypotheses at $j = 2, 3, 4$. Table 3 shows the results for $z_0 = 1/2$, where previously there was no power at the finest scale and therefore no occurrence of aliasing. The empirical power for the $j = 2, 4$ columns is high, due to the spectral power present at those scales. However, for scale $j = 3$, for reasonable sample sizes, the rejection rate becomes consistent with the nominal size. The overall rejection rate at $z_0 = 1/2$ is well controlled, but conservative, with respect to the nominal rate.

Table 4 shows the results for $z_0 = 1/4$, where the power at the finest scale, $j = 1$, is redistributed into coarser scales according to Theorem 2. As before, the null hypothesis is routinely rejected at scales $j = 2, 4$. In contrast to Table 3, Table 4 shows increasing, and eventually high, power at scale $j = 3$. Hence, we would often reject our null hypothesis in favour of the alternative $H_{\mathrm{A}}: S_j(z_0) > 0$ for all $j = 1, \ldots, J^{\dagger}$, and therefore the overall empirical power of the test increases and reaches 65·8% for $b = 128$. In practice we would, for the time being, accept H_0 at $z_0 = 1/2$ and conclude that no aliasing occurred and nor were white noise components present. However, at $z_0 = 1/4$, although we can reject H_0, we cannot conclude that there was definitely aliasing or that white noise components were present, because we have no information on scales for $j > 4$.

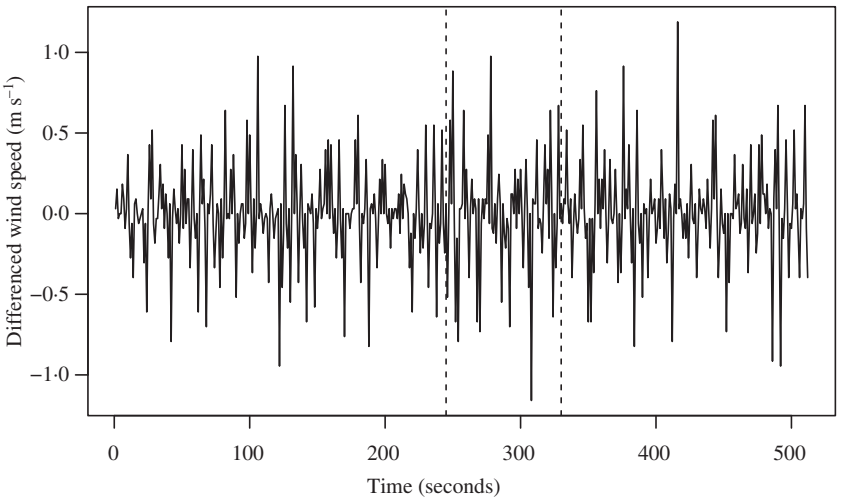

Fig. 2. First differences of wind speed (m s^{-1}) for a proposed wind farm site in the midwestern U.S.A., with $T = 512$. The dashed vertical lines indicate the time-points $t = 245, 330$.

Table 5. *The p-values and H_0 rejection status for the wind series with our test at two reference times, using the signmedian test, three different Daubechies Dv wavelets with v vanishing moments, and $b = 25$*

Time	Wavelet	$j = 1$	$j = 2$	$j = 3$	$j = 4$	Reject H_0
245	D6	0·0	0·3	3·3	0·3	Yes
245	D7	0·0	2·0	6·5	2·0	No
245	D8	0·1	3·3	0·8	0·8	Yes
330	D6	0·0	100·0	0·0	100·0	No
330	D7	0·0	100·0	0·8	100·0	No
330	D8	0·0	100·0	0·1	100·0	No

5·2. *A wind speed example*

Wind power forecasting has received much attention in recent years, motivated by the need to develop reliable forecasting tools to enable effective integration of wind farm output into power grids (Landberg et al., 2003; Genton & Hering, 2007). Several authors have used autoregressive integrated moving average models that implicitly assume the absence of aliasing; see Huang & Chalabi (1995) or Sfetsos (2002), for example. We seek to identify whether any such corruption might occur in data provided by an industrial collaborator. High-resolution wind speed data, sampled at 1 Hz, were acquired from a proposed wind farm in the midwestern U.S.A. during March 2011 and differenced to remove trend. Figure 2 displays the differenced series, which exhibits nonstationarity. For example, the sample variances of the first and last 100 observations are 0·063 and 0·093, respectively, and statistically significant variance differences have been confirmed using methods from Nason (2013). Q-Q plots and goodness-of-fit tests strongly suggest that the series' marginal distribution has heavy tails.

We test H_0 at the two times $t = 245, 330$, and the Holm-adjusted p-values are displayed in Table 5. Results are presented for the three wavelets recommended in §4·2 for the $J^\dagger = 4$ non-cone scales. At time $t = 330$, for each wavelet H_0 is not rejected, so we have no evidence of a white noise component or aliasing at this location. At time $t = 245$, H_0 is rejected at the 5% level

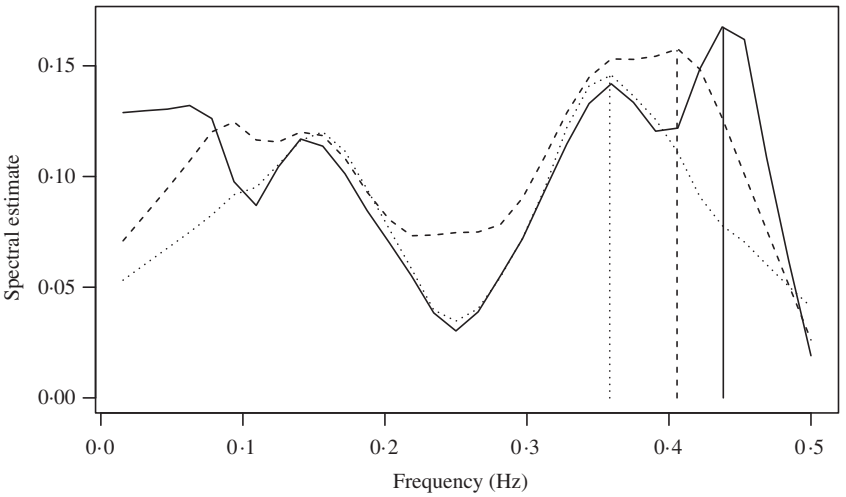

Fig. 3. Three regular local kernel-smoothed periodograms of wind speed series taken over three time windows: 260–323 (solid), 270–333 (dashed) and 280–343 (dotted). The frequencies of the maximum values of each smoothed periodogram are indicated by vertical lines.

for wavelets D6 and D8 but not for D7. However, H_0 would have been rejected at the 6·5% level for the D7 wavelet. Hence there is evidence to reject H_0 in favour of H_A.

We undertake a post hoc investigation as described in §4·3. We did not reject H_0 at $t = 330$, are suspicious that aliasing or white noise components could exist at $t = 245$, and suspect that there might be a transition between the two time-points, possibly around $t = 290$. If prior to $t = 290$ the series was truly subject to aliasing, and after that time it was not, then power will have moved from higher to lower frequencies over time or, at the very least, have disappeared from frequencies above the Nyquist frequency. Since we suspect aliasing prior to $t = 300$, we do not apply classical spectral analysis here.

However, as we have no evidence of aliasing after the suspected transition, we subject the series there to post hoc classical stationary periodogram analyses on three rolling windows, using methods described by Fryzlewicz et al. (2008) and displayed in Fig. 3. The results indicate that the peak frequency associated with each window decreases. Hence, there is supporting evidence for the conjecture that, after $t = 300$, the high-frequency content in the series reduces over time.

Using the centre time, t, for each rolling window and its associated peak frequency, f, we obtain $(t, f) \in \{(292, 0{\cdot}438), (302, 0{\cdot}406), (312, 0{\cdot}395)\}$, which can be modelled approximately by the linear relationship $f = 1{\cdot}06 - 0{\cdot}0022t$. Assuming that the peak frequency reduces linearly according to this model, and extrapolating backwards, we hypothesize that the series was aliased before $t = 261$, when $f = 0{\cdot}5$ in the model. This is merely a hypothesis, as the series could have been contaminated by white noise before the transition rather than by aliased high-frequency components. In addition, the transition time itself, as well as this analysis, is subject to uncertainty.

Acknowledgement

We thank C. Ziesler of Wind Capital Group for the wind speed data and several reviewers who have helped to improve this paper. The authors were partially supported by the Research Councils

UK Energy Programme and gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the proofs of Theorems 2 and 3 and definition of the test-bed spectrum, together with advice on the selection of a suitable hypothesis test of location, and theoretical justification of the asymptotic equivalence of the distribution of $d_{j,k}$ to $d_{j,k+r}$.

APPENDIX

Proof of Theorem 1

First, we substitute $2t$ for t in the main process definition formula (3) to obtain

$$Y_t = X_{2t} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k} \psi_{j,k-2t} \xi_{j,k} = F_t + L_t, \tag{A1}$$

where F_t is the $j = 1$ term of (A1),

$$F_t = \sum_{k=-\infty}^{\infty} w_{1,k} \psi_{1,k-2t} \xi_{1,k}, \tag{A2}$$

and $L_t = V_t + O_t$, with V_t being the $j \geqslant 2$ terms of (A1) corresponding to even-indexed k, i.e., $V_t = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \psi_{j,2\ell-2t} \xi_{j,2\ell}$, and O_t the $j \geqslant 2$ terms corresponding to odd-indexed k, i.e., $O_t = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell+1} \psi_{j,2\ell-2t+1} \xi_{j,2\ell+1}$.

For the next part recall the discrete wavelet formulae in Definition 1. Focusing on the case where $j \geqslant 2$, we concentrate on V_t and O_t. Substituting (2) for $\psi_{j,2\ell-2t}$ in the expression for V_t gives

$$V_t = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \xi_{j,2\ell} \psi_{j,2\ell-2t} = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \xi_{j,2\ell} \sum_{q} h_{2\ell-2(t+q)} \psi_{j-1,q}.$$

Now sum over $r = t + q$ instead of q to obtain

$$V_t = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \xi_{j,2\ell} \sum_{r} h_{2(\ell-r)} \psi_{j-1,r-t},$$

sum over $p = \ell - r$ instead of r to obtain

$$V_t = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \xi_{j,2\ell} \sum_{p} h_{2p} \psi_{j-1,\ell-p-t},$$

and rearrange the terms to obtain

$$V_t = \sum_{p} h_{2p} \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \xi_{j,2\ell} \psi_{j-1,\ell-p-t} = \sum_{p} h_{2p} U_t^{(p)}, \tag{A3}$$

where $U_t^{(p)} = \sum_{j=2}^{\infty} \sum_{\ell \in \mathbb{Z}} w_{j,2\ell} \xi_{j,2\ell} \psi_{j-1,\ell-p-t}$ and we assume that all sums converge absolutely.

In the formula for $U_t^{(p)}$, substitute $q = \ell - p$ to get

$$U_t^{(p)} = \sum_{j=2}^{\infty} \sum_{q \in \mathbb{Z}} w_{j,2(p+q)} \xi_{j,2(p+q)} \psi_{j-1,q-t}.$$

Now define $\xi^*_{j-1,q} = \xi_{j,2(p+q)}$ for $j = 2, 3, \ldots$ and $q \in \mathbb{Z}$; both are sequences of uncorrelated random variables with zero mean and unit variance, and both depend on p. Further, define $W^*_{j-1}(z) = W_j(2z)$ for $j = 2, 3, \ldots$ and $z \in (0, 1)$, and let $w^*_{j-1,q} = w_{j,2q}$. Then we can write

$$U_t^{(p)} = \sum_{j=2}^{\infty} \sum_{q \in \mathbb{Z}} w^*_{j-1,q+p} \psi_{j-1,q-t} \xi^*_{j-1,q} = \sum_{j=1}^{\infty} \sum_{q \in \mathbb{Z}} w^*_{j,q+p} \psi_{j,q-t} \xi^*_{j,q}.$$

Recalling assumption (10) of Nason et al. (2000), we obtain

$$\sup_q \left| w^*_{j,q+p} - W^*_j\{(q+p)/T\} \right| = \sup_q \left| w_{j,2(q+p)} - W_j\{2(q+p)/T\} \right| \leqslant C_j/T.$$

The $\{W^*_j(z)\}$ satisfy the same smoothness conditions as the $\{W_j(z)\}$. Hence $U_t^{(p)}$ is a locally stationary wavelet process for all p. Although the range of p values in (A3) seems to be infinite, only a finite, and typically small, number of h_{2p} are nonzero, so the sum over p and ℓ in (A3) is never infinite.

Therefore, V_t is the sum of a finite number of locally stationary wavelet processes with constant coefficients, not depending on t, and hence is itself a locally stationary wavelet process (Cardinali & Nason, 2010). The same arguments can be applied to O_t. Thus, $V_t + O_t$ is also a locally stationary wavelet process.

Now consider the $j = 1$ term, F_t, from (A2). Clearly, this term has mean zero, as the $\xi_{1,k}$ all have mean zero. Recalling that $\psi_{1,k} = g_k$ from (1), the autocovariance of F_t is

$$\begin{aligned} \operatorname{cov}(F_t, F_{t+\tau}) &= \operatorname{cov}\left\{ \sum_{k=-\infty}^{\infty} w_{1,k} g_{k-2t} \xi_{1,k}, \sum_{\ell=-\infty}^{\infty} w_{1,\ell} g_{\ell-2(t+\tau)} \xi_{1,\ell} \right\} \\ &= \sum_{k,\ell} w_{1,k} g_{k-2t} w_{1,\ell} g_{\ell-2(t+\tau)} \operatorname{cov}(\xi_{1,k}, \xi_{1,\ell}) \\ &= \sum_{k} w^2_{1,k} g_{k-2t} g_{k-2(t+\tau)}, \end{aligned} \tag{A4}$$

because $\operatorname{cov}(\xi_{1,k}, \xi_{1,\ell}) = \delta_{k,\ell}$ by assumption.

If $w^2_{1,k}$ is constant, then $S_1(z)$ is constant for all $z \in (0, 1)$, and hence

$$\operatorname{cov}(F_t, F_{t+\tau}) = w_1^2 \sum_{k=-\infty}^{\infty} g_{k-2t} g_{k-2(t+\tau)} = S_1 \delta_{0,\tau}.$$

The last equality holds because of the orthonormality relations satisfied by the quadrature mirror filter coefficients associated with the underlying mother wavelet; see Burrus et al. (1997, formula (5.28)).

If $S_1(z)$ is not constant for all $z \in (0, 1)$, then substitute $k = \ell + 2t$ into (A4) to obtain

$$\operatorname{cov}(F_t, F_{t+\tau}) = \sum_{\ell} w^2_{\ell+2t} g_\ell g_{\ell-2\tau} \tag{A5}$$

$$= \sum_{\ell} \left\{ W_1^2 \left(\frac{\ell + 2t}{T} \right) + O(T^{-1}) \right\} g_\ell g_{\ell-2\tau} \tag{A6}$$

$$= \sum_{\ell} \left\{ S_1(2t/T) + O(L_1|\ell|/T) \right\} g_\ell g_{\ell-2\tau} + O(T^{-1}) \tag{A7}$$

$$= S_1(2t/T)\delta_{0,\tau} + O(T^{-1}).$$

The transition from (A5) to (A6) is due to formula (10) in Nason et al. (2000), and that from (A6) to (A7) is due to the Lipschitz continuity of $W_1(z)$, with L_1 the associated Lipschitz constant, from Definition 1c in Nason et al. (2000). The remainder terms parallel those found in Nason et al. (2000, Proposition 1) and the proof of Theorem 2.

Proof of Corollary 1

This follows by iteration of Theorem 1 and reference to Theorem 2.

Proof of Theorem 2

We proceed by substituting $2^r t$ for t in the formula for X_t in (3). We assume here that the same wavelet is used for the analysis as for the process construction. The periodogram expectation is

$$\begin{aligned} E(d^2_{\ell,m}) &= E\left\{\left(\sum_t X_{2^r t}\psi_{\ell,m-t}\right)^2\right\} = E\left\{\left(\sum_t\sum_j\sum_k w_{j,k}\psi_{j,k-2^r t}\xi_{j,k}\psi_{\ell,m-t}\right)^2\right\} \\ &= \sum_{j,k,n,p} w_{j,k}w_{n,p}E(\xi_{j,k}\xi_{n,p})\sum_t \psi_{j,k-2^r t}\psi_{\ell,m-t}\sum_s \psi_{n,p-2^r s}\psi_{\ell,m-s} \\ &= \sum_{j,k} w^2_{j,k}\left(\sum_t \psi_{j,k-2^r t}\psi_{\ell,m-t}\right)^2 = \sum_{j,k} w^2_{j,k}P(j,k,\ell,m,r) \quad (\ell\in\mathbb{N};\ m\in\mathbb{Z}), \end{aligned} \tag{A8}$$

where $P(j,k,\ell,m,r) = (\sum_t \psi_{j,k-2^r t}\psi_{\ell,m-t})^2$.

We then substitute $k = n + 2^r m$ into (A8) to obtain

$$\begin{aligned} E(d^2_{\ell,m}) &= \sum_j\sum_n w^2_{j,n+2^r m}P(j,n+2^r m,\ell,m,r) \\ &= \sum_{j,n}\left\{S_j\left(\frac{n+2^r m}{T}\right)+O(T^{-1})\right\}P(j,n+2^r m,\ell,m,r) \\ &= \sum_{j,n}\left\{S_j(2^r m/T)+O(T^{-1})+O(nT^{-1})\right\}P(j,n+2^r m,\ell,m,r) \qquad \text{(A9)} \\ &= \sum_{j,n} S_j(2^r m/T)P(j,n+2^r m,\ell,m,r)+O(T^{-1}) \\ &= \sum_j S_j(2^r m/T)\sum_n P(j,n+2^r m,\ell,m,r)+O(T^{-1}). \qquad \text{(A10)} \end{aligned}$$

The remainders in (A9) for Daubechies' wavelets are derived using standard arguments (Nason et al., 2000, Proposition 4). The Supplementary Material gives details of the remainder for the Shannon case.

Now

$$\begin{aligned} \sum_n P(j,n+2^r m,\ell,m,r) &= \sum_{n,t,s}\psi_{j,n+2^r m-2^r t}\psi_{\ell,m-t}\psi_{j,n+2^r m-2^r s}\psi_{\ell,m-s} \\ &= \sum_{t,s}\psi_{\ell,m-t}\psi_{\ell,m-s}\sum_n \psi_{j,n+2^r(m-t)}\psi_{j,n+2^r(m-s)} \\ &= \sum_{t,s}\psi_{\ell,m-t}\psi_{\ell,m-s}\Psi_j\{2^r(s-t)\}, \end{aligned}$$

where $\Psi_j(\tau)$ is the autocorrelation wavelet from Nason et al. (2000). With $v = s - t$,

$$\sum_n P(j,n+2^r m,\ell,m,r) = \sum_v \Psi_j(2^r v)\sum_t \psi_{\ell,m-t}\psi_{\ell,m-v-t} = \sum_v \Psi_j(2^r v)\Psi_\ell(v).$$

Lemma 1 of Eckley & Nason (2005) shows that when $j > r$ we have $\Psi_j(2^r v) = \Psi_{j-r}(v)$. Further, for $j = r$ we have $\Psi_j(2^r v) = \Psi_1(2v) = \delta_{v,0}$, and for $j < r$ we have $\Psi_1(2^{r-j}v) = \delta_{v,0}$.

Hence, omitting the remainder for the moment, splitting the sum in (A10) at r gives

$$\begin{aligned} E(d^2_{\ell,m}) &= \sum_{j=1}^{r} S_j(2^r m/T) \sum_v \Psi_j(2^r v)\Psi_\ell(v) + \sum_{j=r+1}^{\infty} S_j(2^r m/T) \sum_v \Psi_j(2^r v)\Psi_\ell(v) \\ &= \sum_{j=1}^{r} S_j(2^r m/T) \sum_v \delta_{v,0}\Psi_\ell(v) + \sum_{r+1}^{\infty} S_j(2^r m/T) \sum_v \Psi_{j-r}(v)\Psi_\ell(v) \\ &= \sum_{j=1}^{r} S_j(2^r m/T)\Psi_\ell(0) + \sum_{r+1}^{\infty} S_j(2^r m/T) A_{j-r,\ell} \\ &= \sum_{j=1}^{r} S_j(2^r m/T) + \sum_{j=r+1}^{\infty} A_{j-r,\ell} S_j(2^r m/T) \quad (\ell \in \mathbb{N}; m \in \mathbb{Z}). \end{aligned} \tag{A11}$$

When not subsampling, i.e., $r = 0$, formula (A11) reduces to the formula for the expectation of the raw wavelet periodogram in Nason et al. (2000, Proposition 4).

References

Bloomfield, P. (2000). *Fourier Analysis of Time Series: An Introduction*. New York: Wiley, 2nd ed.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. Philadelphia: SIAM.

Burrus, C. S., Gopinath, R. A. & Guo, H. (1997). *Introduction to Wavelets and Wavelet Transforms: A Primer*. Upper Saddle River, New Jersey: Prentice Hall.

Cardinali, A. & Nason, G. P. (2010). Costationarity of locally stationary time series. *J. Time Ser. Economet.* **2**, Article 1.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*. London: Chapman and Hall/CRC, 6th ed.

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics*, T. Subba Rao, S. Subba Rao & C. Rao, eds., vol. 30. Amsterdam: Elsevier, pp. 351–413.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

de Menezes, L. M., Houllier, M. & Tamvakis, M. (2016). Time-varying convergence in European electricity spot markets and their association with carbon and fuel prices. *Energy Policy* **88**, 613–27.

Eckley, I. A. & Nason, G. P. (2005). Efficient computation of the discrete autocorrelation wavelet inner product matrix. *Statist. Comp.* **15**, 83–92.

Eckley, I. A. & Nason, G. P. (2014). Spectral correction for locally stationary Shannon wavelet processes. *Electron. J. Statist.* **8**, 184–200.

Fryzlewicz, P. (2005). Modelling and forecasting financial log-returns as locally stationary wavelet processes. *J. Appl. Statist.* **32**, 503–28.

Fryzlewicz, P. & Nason, G. P. (2006). Haar–Fisz estimation of evolutionary wavelet spectra. *J. R. Statist. Soc.* B **68**, 611–34.

Fryzlewicz, P., Nason, G. P. & von Sachs, R. (2008). A wavelet-Fisz approach to spectrum estimation. *J. Time Ser. Anal.* **29**, 868–80.

Fryzlewicz, P., Van Bellegem, S. & von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.* **55**, 737–64.

Genton, M. & Hering, A. (2007). Blowing in the wind. *Significance* **4**, 11–4.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

Hannan, E. J. (1960). *Time Series Analysis*. London: Chapman and Hall.

Hinich, M. & Messer, H. (1995). On the principal domain of the discrete bispectrum of a stationary signal. *IEEE Trans. Sig. Proces.* **43**, 2130–4.

Hinich, M. & Wolinsky, M. (1988). A test for aliasing using bispectral analysis. *J. Am. Statist. Assoc.* **83**, 499–502.

Holm, S. (1979). A simple sequentially rejective multiple hypothesis test procedure. *Scand. J. Statist.* **6**, 65–70.

Huang, M. & Chalabi, Z. S. (1995). Use of time-series analysis to model and forecast wind speed. *J. Wind Eng. Ind. Aerodyn.* **56**, 311–22.

Jones, R. H. (1975). Estimating the variance of time averages. *J. Appl. Meteorol.* **14**, 159–63.

Killick, R., Eckley, I. & Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electron. J. Statist.* **7**, 1167–83.

LANDBERG, L., GIEBEL, G., NIELSEN, H. A., NIELSEN, T. & MADSEN, H. (2003). Short-term prediction—an overview. *Wind Energy* **6**, 273–80.

MICHIS, A. A. (2009). Forecasting brand sales with wavelet decompositions of related causal series. *Int. J. Bus. Forecast. Market. Intel.* **1**, 95–110.

NASON, G. P. (2008). *Wavelet Methods in Statistics with R*. New York: Springer.

NASON, G. P. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Statist. Soc.* B **75**, 879–904.

NASON, G. P., VON SACHS, R. & KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc.* B **62**, 271–92.

NOWOTARSKI, J., TOMCZYK, J. & WERON, R. (2013). Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Econ.* **39**, 13–27.

PRIESTLEY, M. B. (1983). *Spectral Analysis and Time Series, Volumes I and II*. London: Academic Press.

SFETSOS, A. (2002). A novel approach for the forecasting of mean hourly wind speed time series. *Renew. Energy* **27**, 163–74.

SPANOS, P. & KOUGIOUMTZOGLOU, I. (2012). Harmonic wavelets based statistical linearization for response evolutionary power spectrum determination. *Prob. Eng. Mech.* **27**, 57–68.

TORRENCE, C. & COMPO, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78.

VAN BELLEGEM, S. & VON SACHS, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Ann. Statist.* **36**, 1879–924.

WINKELMANN, L. (2016). Forward guidance and the predictability of monetary policy: A wavelet-based jump detection approach. *Appl. Statist.* **65**, 299–314.

WUNSCH, C. & GUNN, D. (2003). A densely sampled core and climate variable aliasing. *Geo-Mar. Lett.* **23**, 64–71.

ZWIERS, F. W. & VON STORCH, H. (1995). Taking serial correlation into account in tests of the mean. *J. Climate* **8**, 336–51.

[*Received on* 7 *February* 2014. *Editorial decision on* 4 *May* 2018]